

(2)

Semiannual Technical Summary

AD-A152 329

Wideband Integrated
Voice/Data Technology

30 September 1984

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS



Prepared for the Defense Advanced Research Projects Agency
under Electronic Systems Division Contract F19628-85-C-0002.

Approved for public release; distribution unlimited.

DTIC
ELECTE
APR 11 1985
S B D

DTIC FILE COPY

85 03 22 006

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. This work was sponsored by the Defense Advanced Research Projects Agency under Air Force Contract F19328-85-C-0002. (ARPA Order 3673).

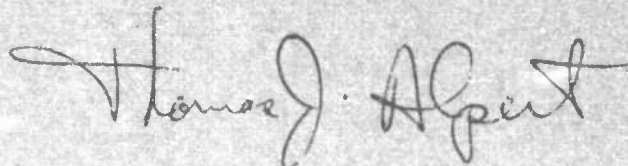
This report may be reproduced to satisfy needs of U.S. Government agencies.

The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.

The Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

A handwritten signature in dark ink, reading "Thomas J. Alpert". The signature is fluid and cursive, with the first name "Thomas" and last name "Alpert" clearly legible.

Thomas J. Alpert, Major, USAF
Chief, ESD Lincoln Laboratory Project Office

Non-Lincoln Recipients

PLEASE DO NOT RETURN

Permission is given to destroy this document
when it is no longer needed.

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY**

WIDEBAND INTEGRATED VOICE/DATA TECHNOLOGY

**SEMIANNUAL TECHNICAL SUMMARY REPORT
TO THE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY**

1 APRIL — 30 SEPTEMBER 1984

ISSUED 30 JANUARY 1985

**DTIC
ELECTE
APR 11 1985
S B**

Approved for public release; distribution unlimited.

LEXINGTON

MASSACHUSETTS

ABSTRACT

This report describes work performed on the Wideband Integrated Voice/Data Technology Program sponsored by the Information Processing Techniques Office of the Defense Advanced Research Projects Agency during the period 1 April through 30 September 1984.

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
Distribution/	
Availability Codes	
Avail and/or	
Dist	Special
A-1	



TABLE OF CONTENTS

Abstract	iii
List of Illustrations	vii
Introduction and Summary	ix
I. SPEECH PROCESSING PERIPHERAL (SPP)	1
II. WIDEBAND EXPERIMENTS AND EXPERIMENT COORDINATION	3
A. Wideband Network System Coordination	3
B. Voice and Data Multiplexing Experiments	5
C. Stream (ST) Flow-Specification Protocol Implementation and Experiments	11
D. Gateway Developments	19
E. SUN Workstations	22
F. Ethernet PVT	22
III. VOICE-CONTROLLED SYSTEMS	25
A. Robust Speech Recognition Overview	25
B. Open-Endpoint Dynamic Time Warping	26
C. DTW-Based Noisy Speech Experiments	28
D. LDSP-Based Noisy Speech Recognition System	30
E. Network-Based Speech Recognition System Development	31
F. Advanced Speech Resource Unit Simulation Facility	33
References	35
Glossary	37

LIST OF ILLUSTRATIONS

Figure No.		Page
1	Compact LPC Speech Processing Peripheral (SPP)	1
2	Voice/Data Integration Experiment Configuration	6
3	Voice/Data Multiplexing in the IP/ST Gateway	6
4	Delay Histogram: Voice/Data Multiplexing Experiment	10
5	(a) ECI CPU Board. (b) ECI I/O Board	23
6	Open-Endpoint Template Matching	27
7	Preliminary Processing Results on AFTI Data	29
8	ASRU System Block Diagram	32
9	DTW Subsystem	32

INTRODUCTION AND SUMMARY

An important challenge in the design of future military communications networks is to achieve system economy and adaptability through efficient and flexible allocation of common network resources to voice and data users. The major objective of the Wideband Integrated Voice/Data Technology Program has been to address this challenge through the development of techniques for integrated voice and data communications in digital packet networks which include wideband common-user satellite links. A major focus of activity in the program is the establishment of an experimental wideband packet satellite network for realistic testing of a variety of strategies for efficient multiplexing of voice and data users. The program also serves as a focus for the development and testing of techniques for local area packet voice distribution, for speech traffic concentration, and for efficient real-time voice communication in an internetwork environment including local networks of various types connected through a wideband demand-assigned satellite network. Beginning in FY84, an additional focus of the program has been on the development of voice-controlled systems, including robust speech-recognition techniques for acoustically noisy environments.

FY84 has been the final year of the Wideband Program effort at Lincoln, and hence this will be the last Semiannual Technical Summary for the Program. Related work on the experimental wideband satellite network will continue in FY85 under the DCA-sponsored Defense Switched Network Technology and Experiments Program. Efforts in robust speech recognition will be transitioned to a new DARPA-sponsored Program in Robust Speech Recognition Technology that commences in FY85. Additional work on wafer-scale VLSI systems for speech recognition will continue in FY85 under the DARPA-sponsored Restructurable VLSI Program.

This report covers work in the following areas: technology transfer of the generalized Linear Predictive Coding (LPC) Speech Processing Peripheral (SPP), coordination and execution of multiuser packet speech and data experiments using the experimental wideband satellite network (WB SATNET), and development of robust speech recognition technology for voice-controlled systems.

The final printed-circuit board version production prototype of the SPP was delivered to Lincoln by Adams-Russell, Inc. (AR), and was tested successfully via speech intelligibility tests and via interoperation tests with the Lincoln prototype. By the end of September, 12 working production units were delivered, with the remaining 48 scheduled for delivery in October along with revised users' manuals. Plans are being coordinated with DARPA to distribute the SPPs to designated contractors and agencies for application in a variety of DARPA and DoD-sponsored programs.

The WB SATNET Task Force effort has continued to address problems of network performance and reliability. A system problem which caused network failure when more than five streams were set up has been identified and diagnosed, and modified PSAT (Packet

Satellite Interface Message Processor) software has been prepared by Bolt, Beranek, and Newman (BBN) to correct the problem. Intensified user-level network tests were run during August and September with the network set in an operational configuration; this provided a number of successful demonstrations of voice and data transmission capabilities, but uncovered some additional system problems. The channel characteristics were improved by a move from the WESTAR III to the WESTAR IV satellite in late July.

A series of voice and data multiplexing experiments has been conducted, exploiting new capabilities in the miniconcentrator gateway to efficiently utilize PSAT streams with multipurpose voice/data packets (MPPs) and gateway-to-gateway (GTG) fragmentation of data packets. Channel utilizations achieved were 90% for integration of Poisson data traffic with voice and 85% for integration of Transmission Control Protocol (TCP) traffic with voice. These results are strongly dependent on the specific mix of voice and data traffic.

An experimental version of a flow specification (Flow-Spec) option for the stream (ST) protocol has been designed and implemented. Flow-Spec, as described in detail in this report, enables the participants in an ST connection (e.g., a packet voice call) to specify their data rate requirements and to negotiate with gateways and networks to obtain the required resources for the connection. The operation of Flow-Spec has been tested successfully in a variety of experimental scenarios, including cases where a calling Packet Voice Terminal (PVT) had (through interactive negotiation with the gateway) reduced its original request for data rate [from 64 kbps Pulse Code Modulation (PCM) to 2.4 kbps LPC] in order to establish a successful connection.

The miniconcentrator gateways have continued to operate reliably. In addition to supporting Flow-Spec, they have been upgraded to provide new conference shutdown capabilities, forwarding to networks in the internet that are not in the gateway's routing table, and slotsize adjustment for voice calls.

Work has continued on integrating packet voice onto a standard Ethernet-type local area network. An Ethernet interface which can be inserted into either a PVT or concentrator interface has been built; final system tests are in progress. SPPs have been interfaced to SUN Microsystems, Inc., workstations with Ethernet capability, and throughput limitations in the SUN UNIX operating system have been uncovered which currently prevent real-time full-duplex LPC voice. Work is in progress in collaboration with SRI International (SRI) and SUN to correct this problem via operating system modifications.

In the area of robust speech recognition, both off-line (VAX-based) and interactive [Lincoln Digital Signal Processor (LDSP)-based] speech recognition facilities have been developed. A new noise-resistant open-endpoint dynamic time warping (DTW) algorithm has been developed and implemented on both the off-line and the interactive systems. The VAX-based system was used to test a new adaptive spectral tilt pre-emphasis technique. For Advanced Fighter Technology Integrator (AFTI) F16 noisy speech tests, adaptive pre-emphasis yielded 2.5 to 5.5 percentage points improvement in recognition accuracy over

fixed pre-emphasis. A new robust spectrally-based front end using adaptive critical band filters developed from a high-resolution spectrum has been implemented on the LDSP, with promising initial results. New network-based recognition structures, including a simulated annealing technique for training Hidden Markov Model (HMM) recognizers, have been implemented and tested on the VAX.

Simulations have been conducted to aid in selecting system parameters for a compact speech recognition system [Advanced Speech Resource Unit (ASRU)] featuring a wafer-scale DTW system being developed in our DARPA-sponsored Restructurable VLSI Program. Attention has focused on a filter bank front end, which allows spectral parameters to be represented with fewer bits than in LPC and which simplifies the internal DTW architecture. System designs for the ASRU have been developed.

I. SPEECH PROCESSING PERIPHERAL (SPP)

The final production prototype of the SPP was delivered by Adams-Russell in April. The unit, depicted in Figure 1, features a flexible LPC-based analysis/synthesis system and an asynchronous RS-232C serial interface for a variety of applications in standalone and host computer environments. The unit, now packaged by Adams-Russell (AR) using a single printed circuit board, comprises about 30 commercial integrated circuits (NMOS, TTL, and analog), and consumes 10 Watts. It was tested satisfactorily both in standalone operation and in conjunction with Lincoln SPP prototypes, and achieved satisfactory Diagnostic Rhyme Test (DRT) intelligibility scores. A 3-speaker Diagnostic Rhyme Test score of 89.5 (standard error 0.47) was obtained for the PC prototype; this score was essentially identical to the scores obtained* with the AR preprototype (wirewrap) unit and with the original Lincoln prototype.

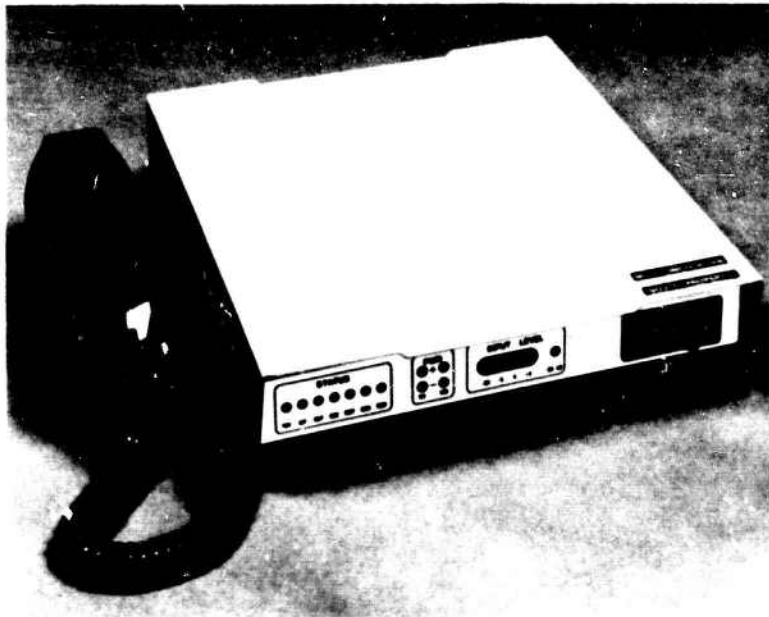


Figure 1. Compact LPC Speech Processing Peripheral (SPP).

Adams-Russell then proceeded with the construction of 60 production units. These production units were identical to the AR prototype except for some minor changes in the RS-232 interface. The first production units were delivered in September. One defect in the PC board was uncovered in the production units. An eyelet on the ground plane was offset so that it occasionally would cause a short between the ground plane and an IC pin. Adams-Russell detected the error and proposed a fix which was submitted to the Lincoln

* Wideband Integrated Voice/Data Technology Semiannual Technical Summary, Lincoln Laboratory, MIT (31 March 1984), DTIC AD-A139924/5, UNCLASSIFIED.

PC Group for evaluation. The overall board quality and the proposed fix were found satisfactory, and all boards were corrected. By the end of September, 12 working production units were delivered, with the remaining 48 scheduled for delivery in October.

Adams-Russell was awarded an additional contract to prepare an updated users' manual for the SPP, to reflect the changes in design and layout between the Lincoln wirewrap prototypes and the production units. A draft users' manual was submitted by AR for Lincoln review. A number of revisions, suggested by Lincoln, will be incorporated in the final manual which will be delivered with the final production units.

Plans are being coordinated with DARPA to distribute the SPPs to DARPA contractors and to other DoD contractors and agencies for use in qualified DoD-sponsored programs requiring flexible speech input/output equipment.

II. WIDEBAND EXPERIMENTS AND EXPERIMENT COORDINATION

A. WIDEBAND NETWORK SYSTEM COORDINATION

The Wideband SATNET Task Force has continued to address problems of network performance and reliability. Gradual improvement has been achieved as problems have been identified and corrected. The most serious system problem that has been attacked in this reporting period is the 'five-stream bug' which was identified in late March. The effect of this bug was to cause all PSATs in the network to crash whenever user traffic levels rose above the limit of five streams. Two Task Force site visits [involving representatives of Lincoln, BBN, Information Sciences Institute (ISI), and LINKABIT Corporation] were carried out in April and July, to address the five-stream bug and other system problems. As discussed below, the specific cause of the five-stream bug was identified in September, and a correction was being prepared for installation in the PSAT at the end of this reporting period. Additional task-force-related activities during this period included: (1) moving the WB SATNET channel to a different satellite (WESTAR IV) with improved channel characteristics in late July; (2) intensified user level tests, with the network set in quasi-operational status, in August and September; and (3) enhanced operational coordination between Western Union and the BBN Network Operations Center (NOC).

A concerted attack on the five-stream bug, and on several lesser network issues, was carried out at a Task Force work session at ISI during the week of 23 April. A set of clues to the nature of the bug were uncovered, which BBN undertook to pursue independently both at BBN and at the Lincoln site. Progress in tracking down the cause of the problem was slowed during May and June by an ESI-A failure at Lincoln which made the Lincoln site inoperative for much of that period. Another Task Force work session was conducted at ISI in July, with emphasis on the five-stream bug and on improving the reliability of remote test and monitoring facilities. An important achievement at this session was development of a technique for precipitating the five-stream bug without crashing every site, thus greatly reducing the turn-around time for each debugging cycle. Subsequent efforts to resolve this problem consisted primarily of extended sessions by BBN personnel using the PSAT at Lincoln whenever possible in the intervals between the operational periods discussed below. The specific cause of the bug was identified by BBN in late September; it involved data being inadvertently written on top of an area of PSAT program memory when an attempt was made to set up more than four streams. The correction required BBN to carry out some restructuring of the PSAT software. By the end of this reporting period, the revised PSAT code had been written and was ready for test.

The five-stream bug caused no problems when only four or fewer sites had the capability to create streams on the net, and therefore an interim policy was established of disabling stream capability (by means of PSAT software patches) at all but four sites at any given time. This policy was administered by BBN, and arrangements were made for particular sites to obtain network access by contacting BBN prior to scheduled

demonstrations or experiments. An example of such an exercise was a demonstration of simultaneous PCM calls to Fort Monmouth and Fort Huachuca from Defense Communications Engineering Center (DCEC) via the Packet/Circuit Interface (PCI) equipment, carried out during an Experimental Integrated Switched Network (EISN) Steering Group meeting at DCEC on 11 April.

With the goal of achieving improved performance for the wideband satellite channel, an agreement was reached between Western Union and the sponsors to transfer the net to a different satellite capable of providing increased signal-to-noise ratio. The transfer was accomplished for all seven network sites during a brief period in late July and early August, with Lincoln providing power and frequency calibration support for Western Union. The new parameters are: uplink frequency 6013 MHz, downlink frequency 3788 MHz, on Transponder 2-Cross on WESTAR IV, at 99 degrees west longitude. The corresponding old parameters were 5959 MHz up, 3734 MHz down, Transponder 1 on WESTAR III, 91 degrees W. In August, Lincoln supported Western Union in calibrating power and frequency for the eighth and ninth sites on the network, namely LINKABIT and Carnegie-Mellon.

On 9 August, the Wideband Network Task Force met at BBN (Cambridge) to present a progress and status report to DCA and DARPA. The primary focus of the meeting was to address issues concerning achievement of stable, routine operational status for the network. Although several problems were known to exist (such as the five-stream bug), it was apparent that there was no fundamental impediment to quasi-operational use of the network on at least a part-time basis. It was decided, therefore, that the network would be made available for user traffic each Thursday and Friday until further notice, with debugging and maintenance confined to the other three days of the week. Pending the correction of the five-stream bug, a limit of four was placed on the number of sites capable of generating streams simultaneously, as noted above. Arrangements were made for any sites having particular need for network access at a specified time to arrange for it by contacting BBN.

Quasi-operational availability as defined above was put into effect on Thursday, 16 August. The primary user voice traffic was provided via access from the telephone network to PVTs with ISI Switched Telephone Network Interface (STNI) cards on Lincoln Experimental Packet Voice Network (LEXNETs) at ISI, SRI, DCEC and Lincoln. All of these were made available to the user community by distributing telephone numbers and instructions for access from the public telephone network. BBN provided the necessary manpower to supervise the net each Thursday and Friday, and produced a report at the end of each week giving details of the operation experience. While many calls were made successfully, the concentrated operational tests disclosed a number of system problems in addition to the five-stream bug mentioned above. These included: (1) occasional halts of the ESI/PSAT global time clocking, (2) inability to set up certain configurations of multisite calls and conferences, and (3) unsatisfactory subsystem reliability. Concentrated efforts to resolve these problems were in progress at the end of the reporting period.

On 11 September, a meeting was held at Western Union's WESTAR Control Center at Glenwood, New Jersey, to clarify and enhance operational coordination between Western Union and the BBN Network Operations Center. The attendees included representatives of Western Union, BBN, and Lincoln Laboratory. The results of the meeting were:

- (1) it was agreed that BBN NOC would report all Western Union troubles to Glenwood,
- (2) WB SATNET site personnel could contact local Western Union repair centers directly if desired, for convenience, provided that BBN NOC was contacted also, and
- (3) the DAQ, Inc., remote monitoring/control equipment was successfully demonstrated during a tour of the Glenwood facilities.

B. VOICE AND DATA MULTIPLEXING EXPERIMENTS

A goal of our work in the Wideband Program has been to explore the effectiveness of packet techniques for multiplexing voice and data traffic. In particular, our work has focused on the use of the WB SATNET for carrying the voice and data traffic. In this section, we describe the voice/data multiplexing that takes place in our IP/ST gateways and present our conclusions from experiments undertaken to evaluate that capability. (The notation, IP/ST, indicates that our gateways accommodate both the DoD standard Internet datagram Protocol and the Lincoln-developed stream or ST protocol, a virtual-circuit type of protocol designed for efficient handling of real-time traffic such as voice.)

In order to support voice effectively, a packet system must provide enough channel capacity to handle the peak voice encoding rate during talkspurts, and must have a predictable packet delay dispersion so that talkspurts can be reconstituted smoothly at the receiver's voice terminal from packets that arrive at somewhat irregular intervals. The wideband satellite network (WB SATNET) offers a 'stream' service that is well suited for voice transmission. SATNET streams are reservations of uplink channel capacity that allow gateways to send packets at prearranged periodic intervals. The capacity of a stream reservation is set by negotiation between a gateway and the SATNET and can be adjusted to handle the encoding rate requirements of the voice calls to be handled. The general strategy is to reserve enough stream capacity to handle the voice load on the basis of information provided in the Flow-Spec parameters in the ST protocol messages used to set up the voice connections.

Because of the talkspurt nature of voice traffic, and the fact that packet voice systems need not transmit packets during the silent intervals between talkspurts, a voice call will, on the average, utilize only about one half of the stream capacity needed to sustain the peak rate during talkspurts. The other half can be used either to squeeze in more voice traffic and/or to carry data traffic. The statistics of voice traffic are such that efficient use of overall capacity for voice alone does not occur unless the number of talkers being

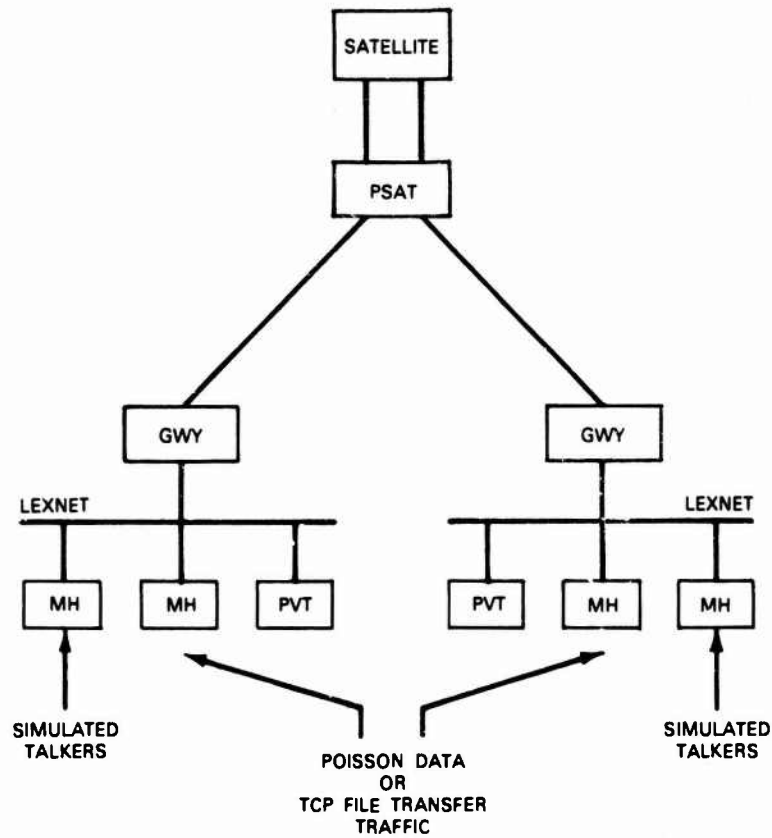


Figure 2. Voice/data integration experiment configuration.

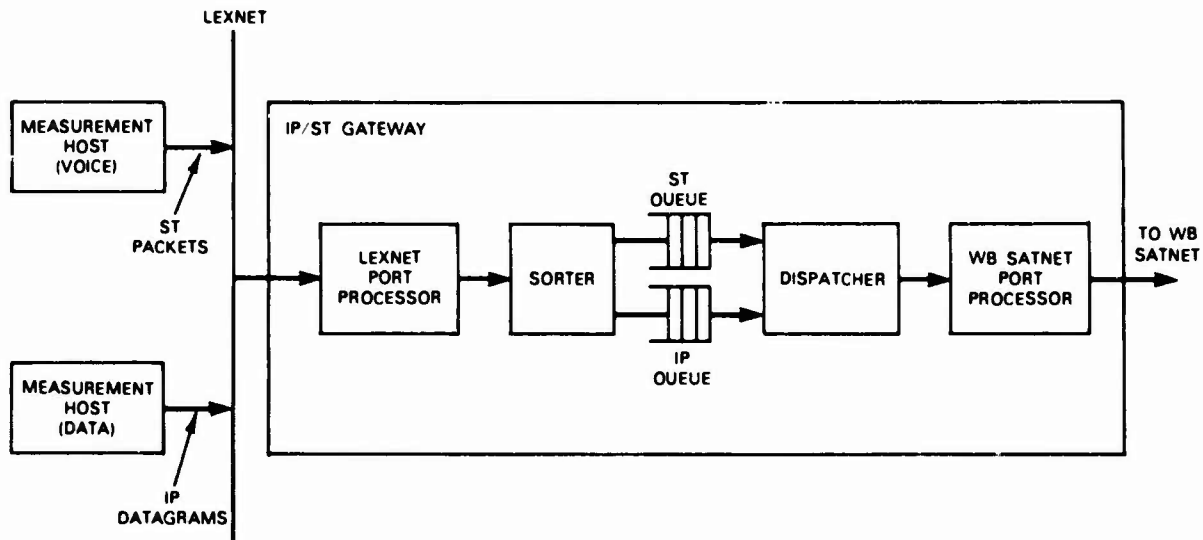


Figure 3. Voice/data multiplexing in the IP/ST gateway.

multiplexed is relatively large (25-50). For networks with capacity for just a few simultaneous talkers, the leftover capacity can be utilized only by traffic that can tolerate rather highly variable delay characteristics. Our recent work on voice/data multiplexing has been directed toward achieving efficient capacity utilization in such situations.

It should be noted that the WB SATNET offers another type of service called 'datagram' service that is intended to make use of uplink channel capacity that is not reserved for streams. In principle, this service would be ideal for handling cases where the offered data traffic exceeded the leftover stream capacity. Unfortunately, current implementation of this service in the PSATs lacks protection against overloads, and its use causes the PSATs to crash at exactly the time that datagram service would be most helpful. As a result, we have done all our experimental work using stream service only.

Our explorations of voice/data multiplexing have primarily made use of configurations such as that shown in Figure 2. We use Measurement Hosts (MHs) to generate and measure simulated voice and data traffic, and Packet Voice Terminals (PVTs) to place real voice calls against a background of simulated traffic. The MHs are PVTs with a special timer card to get precise timing for packet generation rates and delay measurements. Software for the MHs can generate IP datagrams with deterministic or Poisson statistics or ST voice packets with a multitalker statistical talkspurt model. Other MH software can simulate a file transfer using a TCP/IP protocol (the DoD standard Transmission Control Protocol and Internet Protocol) model. Measurements include counts of packets sent, received, retransmitted, etc., as well as delay histograms. Instrumentation in the gateways provides additional information about traffic characteristics.

Figure 3 shows a simplified functional diagram of the voice/data multiplexing that occurs as packets flow from their sources in MHs on a LEXNET to the WB SATNET. Multiplexing functions take place at several points on this diagram. The first is on the LEXNET cable as packets contend for access to the LEXNET port processor on the gateway. The interface between the LEXNET and the gateway has an upper limit on packet handling capacity. We have two options for dealing with burst peaks that equal or exceed this limit. One is to drop the packets that cannot be handled at the instant they arrive. The other is to force a collision indication on the LEXNET that will cause the sender to retransmit after a timeout, thereby introducing a backpressure relative to the packet stream. The backpressure prevents packet loss during peak load periods at the expense of increased delay. In either case, contention at the gateway interface affects voice and data traffic equally and does not cause packet loss or significant delay except at average traffic loads that approach the limit of port capacity. In normal gateway operation, we use the loss option.

Once in the Port Processor, the packets must contend for memory resources in the gateway. Here, our standard strategy is to give the voice packets some preference. If the supply of packet buffers falls below a threshold, only voice packets will be accepted. The rationale for dropping data packets in favor of voice is based on the assumption that higher

level protocols such as TCP can and will retransmit the dropped packets while such a recovery mechanism is not available for the voice traffic. Again, this contention mechanism will cause data packet loss only at high average load where overall gateway resources are being stressed.

The Port Processor hands the packets over to the Sorter which has the function of routing the packets and putting them on queues for dispatching to the output networks. Dispatching is done using a clock to assure timely departure of voice traffic so that the controlled overall delay variance requirements of that type of traffic can be met. In the case of the WB SATNET, the clock rate is adjusted to be a little slower than the rate at which stream packets are sent to the satellite to make sure that packets will not accumulate in the sending PSAT. Voice and data are queued separately, and different disciplines are used for limiting the queues. Our discipline for the voice queue is to allow a packet to remain on the queue only for a limited number of dispatch intervals (typically three). If the stream size is properly adjusted for the voice load, we do not observe packet loss with this discipline. For the data queue, we set a limit on the total number of bytes represented by the IP datagrams on the queue. An arriving packet that finds the queue over its limit will be dropped, and an Internet Control Message Protocol (ICMP) SOURCE QUENCH message will be generated to be dispatched to the sender of the dropped packet.

The Dispatcher operates in a network-dependent fashion to deal with the particular properties of individual networks. For the WB SATNET, the Dispatcher builds multipurpose packets (MPPs) that combine ST voice packets, IP data packets, and special gateway-to-gateway (GTG) fragments of IP datagrams that are all destined for the same other gateway on the WB SATNET. We use these MPPs to get more traffic through the WB SATNET than we could get if the original packets were transmitted directly through that net. The advantage comes because, though the WB SATNET has a rather severe limit on the number of packets it can handle per unit time, the packets can be large in comparison to the original voice packets and to many typical-sized data packets. By combining many small packets into a few large MPPs we can achieve significantly higher total throughput.

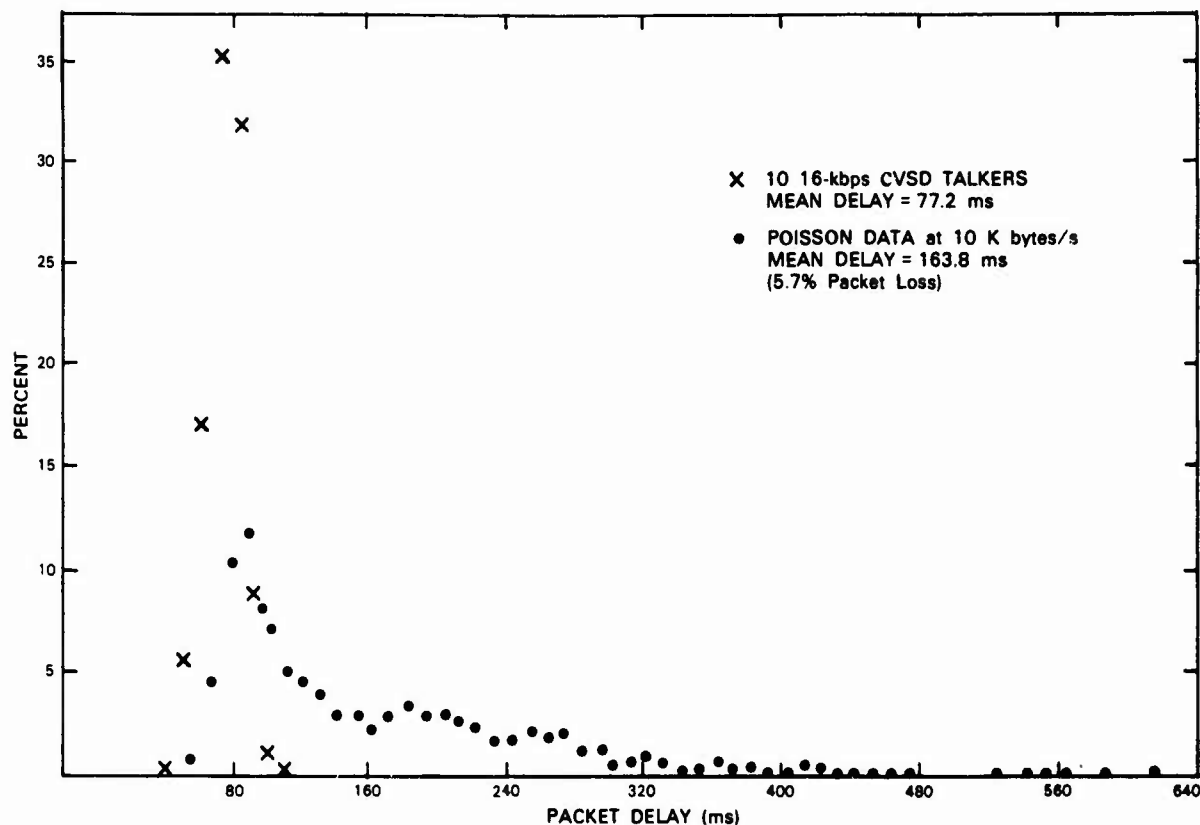
To handle the building of MPPs, the Dispatcher maintains voice and data queues for each destination gateway (site) to which it might have traffic. (We show only one pair of such queues in the simplified diagram of Figure 3.) When the clock signal occurs, the Dispatcher calculates the total quantity of voice traffic on all the site queues to determine how much, if any, data traffic can be dispatched in the upcoming interval. After this planning pass through the queues, the Dispatcher goes around again building MPPs. It gives priority to the voice packets, and fills the remaining stream capacity with IP datagrams and/or GTG fragments until either the slot fills or the queues are exhausted. Round-robin algorithms are used to maintain fairness during periods when queued traffic exceeds channel capacity.

The use of GTG fragments allows efficient filling of the stream slot without regard to the sizes of the IP datagrams on the data queue. Without some kind of fragmentation, IP

datagrams larger than the slot size could never be sent, and others might wait long periods for a slot with enough leftover capacity. The IP protocol defines a fragmentation mechanism that could be used in this situation, but IP fragments have a relatively high header overhead (20 bytes per fragment compared to 8 for GTG fragments), and since they are reassembled only at the destination host, their use would increase the number of packets that downstream networks would have to handle. Further, the reassembly of IP fragments is a complex process since fragments may arrive out of order, and there is no straightforward way to decide when to stop waiting when a missing fragment is detected. In contrast, the reassembly of GTG fragments is relatively simple since the fragments arrive in order, and the failure of a fragment to arrive in its expected order is a clear signal that the reassembly can be aborted. Our implementation does not attempt to provide reliable delivery of reassembled IP datagrams by using any sort of acknowledgement/retransmission mechanism. This policy is consistent with the definition of IP datagram service.

A series of experiments has been carried out to test the effectiveness of the mechanisms just described. The experimental configuration is shown in Figure 2. Most experiment runs made use of the direct connection between the gateways since the path through the PSAT and the satellite was not always available, and we observed no effect other than overall delay in going through the satellite. Two of the MHs (one on each LEXNET) emulated and measured the voice traffic. Parameters for the voice talkspurt model were set for ten 16-kbps (CVSD rate) conversations with packets containing 40 ms of speech transmitted only during talkspurts. The other two MHs generated and measured data traffic that was varied in a series of experiment runs to explore the extent to which the leftover capacity could be utilized. In some runs, a Poisson traffic model was used to measure data delay as a function of offered load. In others, a TCP file transfer model was used to determine the rate at which a single file transfer could proceed through the time-varying residual channel. The capacity of the link connecting the gateways was set to a value roughly twice the average voice requirement.

Overall, the results showed that the gateway multiplexing mechanisms worked as intended. Voice traffic received the desired preferential treatment with throughput and delay almost unaffected by the level of contending data traffic. With overall traffic levels below saturation of the gateway port to the LEXNET, we observed only a small increase in mean voice packet delay and dispersion as data traffic increased. At the same time, we observed a marked increase in mean data delay and dispersion as well as data packet loss as channel capacity limits were approached. Figure 4 shows histograms of measured delays for a typical experiment run in which the mean data traffic was equal to the mean voice traffic load. Overall offered traffic for this run was about 90% of channel capacity. The plot shows a rather tight and symmetrical distribution for the voice packet delays and a skewed distribution with a long tail plus some lost packets for the data. For really high traffic loads that pushed the LEXNET port into saturation, we observed voice packet losses as well. The ST protocol is designed to avoid such load conditions by refusing connections that would saturate the communication capacities. Unfortunately, IP lacks any mechanism for dealing with such a situation, and an overload of data traffic can thus spoil packet voice communications.



147578-N

Figure 4. Delay histogram: voice/data multiplexing experiment.

The results from experiments using the Poisson data traffic showed that the gateway-to-gateway fragmentation and reassembly capability allowed about a 90% utilization of the link capacity with data losses of a few percent and acceptable values for mean data delays and delay dispersions. The best results obtained with the TCP file transfer showed about 85% utilization. Utilization was lower in this case because the packet losses caused retransmissions, and some transmission opportunities were missed waiting for timeouts. This performance is much better than we achieved in earlier experiments without fragmentation, but a quantitative comparison is not meaningful because the performance without fragmentation depends very strongly on the particular choice of data packet size in relation to voice packet size and channel slot size.

Channel utilization could be improved by increasing the size of the Dispatcher's data queue at the expense of a corresponding increase in average delay and delay variance. It is clear from both queueing theory and experience that the price in delay increases rapidly beyond the 85-90% region of average utilization that we have been exploring experimentally. Excess buffer capacity is likely to be detrimental to overall performance because it can result in large delay variances in situations such as ours where the effective channel capacity

for the class of traffic varies rapidly. Protocols such as TCP that use a measurement of round-trip delay to set retransmission timeout values perform poorly in situations in which high delay variances are encountered. They will tend to retransmit excessively if the estimated delay is lower than the actual, or waste opportunities if the estimate is too long. By limiting queue lengths (ideally on a connection basis, for fairness) network nodes can reduce delay variance and improve overall performance for such protocols.

It should be noted that the values for channel utilization obtained in these experiments are strongly dependent on the particular combination of voice traffic load and link capacity chosen for the experiments. If the number of talkers had been larger, and the link capacity correspondingly increased, we would expect to achieve better utilization of the residual capacity because there would be less variation in the voice load from moment to moment. Similarly, fewer talkers in a smaller channel would lead to greater variability and less success in utilizing the residual capacity.

An often misunderstood aspect of WB SATNET stream service is that the reservation of uplink capacity is not dedicated to any particular destination on the downlink. If a gateway has traffic for three other gateways on the net, it does not ask for three streams, it merely increases the size of its one stream to handle the expected traffic. As a result of this pooling of resources, the possibilities for efficient statistical multiplexing are much better than they would be if it were necessary to reserve a stream for each destination.

C. STREAM (ST) FLOW-SPECIFICATION PROTOCOL IMPLEMENTATION AND EXPERIMENTS

We have designed and implemented an experimental version of a Flow-Specification (Flow-Spec) option for ST protocol. In this section, we provide a motivation for Flow-Spec and describe its fields and its implementation for point-to-point connections in the ST Gateway and in the LEXNET PVT terminals. We then describe experimental scenarios which show the operation and negotiation possibilities of Flow-Spec. Finally, we describe the simpler Flow-Spec for speech conferencing.

1. Motivation for ST Flow-Spec

One of the goals of the ST protocol and its implementation in the IP/ST Gateway (hereafter referred to as 'Gateway') is the timely transmission of speech. To achieve this, the Gateway reserves network capacity that is adequate for the anticipated load when a connection opens. Since network capacity can be an expensive resource, e.g., stream slotsize on the WB SATNET, it is incumbent on the Gateway to match allocated capacity with the actual needs. If network resources become overallocated, the Gateway needs to know which connections can adapt to decreased demands. This suggests that the parties involved in a connection inform the Gateway of their needs and constraints so that the Gateway can act accordingly.

Flow-Spec is that part of the ST protocol that enables the participants in an ST connection to specify various capacity requirements and limitations relating to the connection. Using this information, the participating Gateway(s) can determine the availability of resources for the connection and can negotiate, where appropriate, with the end users for lower demands if the requested capacity is not available at connect time or becomes unavailable later on.

As currently specified, the parameters in a Flow-Spec are oriented towards packet speech communication. However, we believe that the presently defined parameters can serve for other applications, such as packet video, that involve more or less constant flow rates for periods that are long in relation to packet transmission times. They are not suitable for normal data communication applications, such as file transfers or interactive terminal support. Two types of Flow-Spec have been defined, one for point-to-point connections and the other for conference connections. Since Flow-Spec for conference connections is a subset of that for point-to-point connections, we first describe Flow-Spec for point-to-point connections.

2. Flow-Spec for Point-to-Point Connections

The initiator of a connection (the 'ORIGIN') describes the proposed data flow specifications of the connection by appending a Flow-Spec to a CONNECT message. The Flow-Spec contains two sets of nine parameters each, one set describing the characteristics of the connection in the 'forward' direction (i.e., forward from the ORIGIN) and the other set describing the 'backward' direction (i.e., backward to ORIGIN). As the Flow-Spec travels from one Gateway to the next one, it is examined and some of the parameters may be modified to reflect additional limitations. A Gateway along the way may send a REFUSE to reject a proposed connection which is deemed to require resources that cannot be supported. Eventually, the CONNECT reaches the end user (the 'TARGET') who may accept the connection (with or without further limitations) or may refuse it.

In the general case, the flow requirements of a proposed connection should have an impact on the choice of a route for the connection. Because our Gateways do not operate in an environment with interesting alternate route possibilities, we have not implemented alternate routing capabilities. At any one time, our Gateways have only one possible route to explore in getting to a particular destination. We have therefore not explored the known-to-be-difficult problem of routing under the multiple constraints posed by the Flow-Spec.

The nine parameters in the Flow-Spec are divided logically into two categories, the request and the response. The request parameters are generated by the ORIGIN to specify the desired characteristics of the proposed connection. They are:

- (1) Packet size (number of words of data in each packet)
- (2) Peak packet rate (10 multiplied by the number of packets per second)

- (3) Duty cycle (percent between 0 and 100 indicating the fraction of time that packets will be generated at the indicated peak rate)
- (4) Negotiability bits (setting of one of the three bits a-c indicates that the corresponding quantity is at a minimum and therefore is 'non-negotiable')
 - (a) Proposed packet size is minimum that can support the end user's needs ('MIN-SIZE')
 - (b) Proposed peak packet rate is minimum that can support the end user's needs ('MIN-RATE')
 - (c) Proposed capacity (peak packet rate multiplied by packet size) is minimum that can support the end user's needs ('MIN-PRODUCT')

The response parameters are initialized by the ORIGIN. They are updated by the ST agents along the path from the ORIGIN to the TARGET as the CONNECT is propagated in the forward direction and as the ACCEPT returns in the backward direction. They are:

- (1) Limitation bits
 - (a) Proposed packet size has been determined by an ST agent in the path to be too big
 - (b) Proposed peak packet rate has been determined by an ST agent in the path to be too big
 - (c) Proposed capacity has been determined by an ST agent in the path to be too big
- (2) Maximum size packet that can be supported by the network(s) in the path. (This value is decreased, as necessary, as the Flow-Spec travels from one Gateway to another, thereby reflecting the maximum size that can be supported by the most limiting network in the path)
- (3) Maximum peak packet rate that can be supported on the network(s) in the path. (Like the maximum size parameter, this is decreased as necessary to reflect the maximum rate of the most limiting network in the path)
- (4) Predicted mean delay for a packet traveling on the connection from its source to this point (number of milliseconds)
- (5) Predicted delay variance for a packet traveling on the connection from its source to this point (number of square milliseconds). This parameter and the previous one are intended to provide advisory information to end users for planning buffering and speech reconstitution strategies. They represent predictions based on *a priori* knowledge about the delay characteristics of the networks and do not imply any 'on-line' measurement capabilities in the Gateways or guarantees of service meeting the predicted delay characteristics.

3. Gateway Strategy

A typical scenario for the opening of an ST connection begins with the ORIGIN issuing a CONNECT to its neighboring Gateway. (Presumably, at an earlier stage, the ORIGIN and TARGET decided upon basic requirements and capabilities of both ends by using a high-level protocol, e.g., NVP for speech communication, but that is of no direct concern here.) The CONNECT has a Flow-Spec which contains, for both directions, a set of request parameters and an initialized set of response parameters. The Gateway examines the Flow-Spec parameters and modifies the response parameters where necessary. Limitation bits are set, if warranted. Then the Gateway passes the CONNECT on to the next Gateway which incorporates still further responses, as appropriate. This continues until the TARGET receives the CONNECT. If, at any stage in the transmission path, the resulting combination of requests and responses present a situation that is not viable, then the Gateway sends a REFUSE to reject the proposed connection on the spot. Assuming the CONNECT was not refused along the way, the TARGET then responds with an ACCEPT. (The TARGET could, in principle, further limit the proposal or even REJECT it, but usually simply agrees to it.) The ACCEPT then travels all the way back to the ORIGIN. If the ACCEPT has any limitation bits set, then the ORIGIN must now issue a DISCONNECT or a new CONNECT with a new set of requests if other flow characteristics would be useful. In the latter case, the cycle starts all over.

In order to avoid the possibility of Gateways needing to 'serve two masters,' we have imposed the restriction that only the ORIGIN can request the packet size, peak packet rate, duty cycle, and negotiability bits for both the forward and backward directions. To enforce this unidirectionality of Flow-Spec requests, each Gateway maintains in its data base the Flow-Spec requested for a connection and checks arriving ACCEPTs for matches of the packet size, peak packet rate, duty cycle, and negotiability bits. If a nonmatch is encountered, then the Gateway discards the ACCEPT as being out-of-date, the assumption being that the ACCEPT was generated in response to an earlier CONNECT that has since been superseded.

When a Gateway has a proposed Flow-Spec that does not have limitation bits set, it attempts to allocate network resources based on the requested parameters. The success or failure of the attempted allocation is determinable in two stages, immediate and delayed. An immediate conclusion can be reached for those networks where the Gateway itself manages its capacity on the network and does not have to request resources from another party. An example of such a network is LEXNET. Additionally, failure to allocate the resources can be determined immediately for any network if the Gateway notices that the cumulative resources for the new connection in conjunction with the resources already allocated exceed the maximum resources the Gateway knows can be supported on that network. The delayed conclusion is reached in the case of the WB SATNET where the Gateway must request from its PSAT an enlargement to the stream slotsize to accommodate the new connection and must wait for a response from the PSAT to this request. Pending the PSAT response, the Gateway tentatively assumes that the enlargement will be granted and proceeds accordingly.

If a request to enlarge the stream slotsize ultimately is denied by the PSAT, then the Gateway is over-committed in the resources it has tentatively allocated and must, therefore, decrease one of its commitments. Since the response from the PSAT is asynchronous with other Gateway activities, including the possible opening and closing of other connections, a general approach was taken in the design of the Gateway reaction to the denial of the stream enlargement request. The Gateway searches in a priority fashion for possible candidates that could yield some resources. First, it searches its data base for connections that have indicated negotiability for their requested capacity (packet size multiplied by peak packet rate). If any such connections are found, then the most recent one is requested to decrease its packet size by sending to the ORIGIN of the connection a new ACCEPT with the limitation bit set indicating that the packet size is too big. If no such connections are found then a search is made for connections that have indicated negotiability in packet size. Finally, as a last resort, the most recent connection is closed down via a DISCONNECT.

Our current implementation treats the allocation of resources due to a Flow-Spec in a one-dimensional fashion. For example, to accommodate a new connection, we enlarge the WB SATNET stream by an amount equal to the requested packet size. A more general allocation strategy would take into account the packet sizes, peak packet rates, and duty cycles of connections already allocated in order to determine a likely worst case of packets buffered and waiting to be transmitted. Allowing various buffering strategies and packet packing combinations, we then could decide how much of a stream increase really was needed to accommodate the new connection. In fact, we might not have to increase the stream at all to accommodate a connection whose packet size and rate would permit its packets to 'slip in' to frames that were only partially full.

4. PVT Flow-Spec Implementation

A goal of our work has been to demonstrate the ability of packet voice techniques to adapt to changing network resources by switching encoding rates both at call setup time and dynamically during a call. To support such a demonstration, the Lincoln PVT is designed to have a PCM codec as standard equipment plus space for an optional narrow-band vocoder card. The design allows software in the PVT to sense the presence or absence of the optional card, and, if it is present, to determine the vocoder type and to select between PCM and the vocoder. A momentary switch on the PVT allows the user to indicate a preference for PCM or the vocoder. The actual choice of encoding rate for a point-to-point conversation depends on a combination of user preference, the availability of matching encoders at both ends, and a match between the flow requirements of the encoders and the network resources available at the moment. Protocol negotiations at two levels are involved in the process. First, a high level Network Voice Protocol (NVP) negotiation takes place at call setup time to determine that the caller and callee have one or more compatible encoders. Then a negotiation takes place between the calling PVT (the ORIGIN) and the network of Gateways that lie on the path between the ORIGIN and the TARGET (the called PVT) using the Flow-Spec parameters on the CONNECT and ACCEPT messages that are used to set up the ST connection for the call.

In the Flow-Spec negotiation, the ORIGIN requests a flow and the Gateways provide a 'go/no-go' response plus some information about absolute limits on packet sizes and peak packet rates. The ORIGIN follows a trial-and-error strategy to find a flow request that is acceptable to the Gateways and compatible with the available encoders. The TARGET watches the arriving Flow-Spec parameters and switches its encoders to match the accepted flow rates, but does not directly participate in the negotiation with the Gateways. If the user at the TARGET PVT indicates a desire to change the encoding rate by moving the vocoder selection switch, an NVP token is sent to the ORIGIN requesting a change in the Flow-Spec.

Let us consider a scenario in which a call is set up across the WB SATNET between two PVTs each with LPC vocoders (2.4 kbps) plus the standard PCM (64 kbps). The initial NVP communication using IP datagrams determines that both parties have matching encoders and that the TARGET phone is not busy. Success at this point starts the TARGET phone ringing. At the same time, the ORIGIN starts the ST connection setup process by sending a CONNECT message to its local Gateway. Assuming the caller had indicated a preference for PCM, the Flow-Spec on the initial CONNECT would request the packet size and peak packet rate appropriate for PCM encoding in both directions on the connection. The requirement that both directions use the same encoding comes from hardware considerations in the PVTs rather than any limitations in the protocol. The Flow-Spec would have negotiability bits indicating that the requested flow was not at a minimum product of packet size and rate but that the packet size and peak rate were each individually at their minimum values. This apparent contradiction between the MIN-PRODUCT bit and the individual MIN-RATE and MIN-SIZE bits results from the discontinuities associated with the availability of different encoding rates. In this case, the indication that the request is not at minimum product tells the Gateways that some other lower encoding rate is available to the PVT. The indication that the packet size is at minimum and that the peak packet rate is at minimum tells the Gateways that the PVT cannot carry out any trade-offs between packet size and rate for this encoding.

When the first Gateway receives the CONNECT message, it sends a message to the PSAT requesting an increase in the Gateway's allocation of SATNET stream capacity by an amount determined from the Flow-Spec request for the forward direction of the connection. At the same time, it propagates the CONNECT to the second Gateway, which in turn asks for PSAT capacity allocation for the backward direction of the connection and propagates the CONNECT to the TARGET. The TARGET responds with an ACCEPT message that propagates back to the ORIGIN. It is likely that the ACCEPT will get back to the ORIGIN before the PSATs have responded to the stream change requests, and that the limitation bits in the Flow-Spec returned by the Gateways will indicate that the flow can be handled successfully. However, if either PSAT denies the increase in stream allocation, the flow cannot be handled, and the Gateway so informed will generate a new ACCEPT message with limitation bits indicating that the requested packet size is too large. (Really, it is the product of packet size and peak rate that is too large, but the Gateways see the problem as an inability to increase the stream slotsize to handle the requested packet size.)

In the event that the PCM encoding rate cannot be handled, the ORIGIN will attempt to switch the call to LPC encoding to see if the lower rate can be handled. The switch involves sending an NVP message to the TARGET informing it of the intended switch as well as sending a new CONNECT to the Gateway with LPC values for the Flow-Spec. This Flow-Spec will request a peak packet rate of 25 packets per second and a packet size of eight words, which corresponds to two 20-ms parcels of encoded speech plus the NVP header. The negotiability bits will indicate that the request is at minimum product because the PVT has no lower encoding rate and that the requested packet size is at a minimum because the PVT is not prepared to send packets at a higher peak rate which would be required if the size were to be reduced. This Flow-Spec tells the Gateways that the ORIGIN is still able to negotiate a trade-off between size and rate while keeping the product more or less constant, but that the negotiation can only be in the direction of larger packets and lower peak rates. (In such a trade-off negotiation, the product remains only approximately constant since the NVP header overhead stays at two words per packet causing the product to decrease slightly as the packet size increases.)

If the LPC encoding rate can be handled by the network, the actual switch to LPC encoding will take place independently at the two ends on the arrival of CONNECT or ACCEPT messages with limitation bits indicating that the requested flow is acceptable. At the time of switching, there may be some speech packets in flight that will arrive and be processed with the wrong decoder, but experience shows that the resulting glitches, which last only a fraction of a second, do not seriously degrade the communication link. There is also a possibility that voice packets will be sent during periods when the stream capacity for them has not been successfully obtained from the PSATs. In our present implementation, such packets will contend for the allocated capacity and will, with some probability, cause glitches due to lost packets in other established conversations. A more complex implementation might assign packets for connections with unsatisfied flow requirements to a secondary queue so that they would get channel capacity only when all established connections had been served. Our experience suggests that the added complexity probably is not needed since the duration of such glitch intervals is quite short. In the simple case of an attempted PCM call switching to LPC at the start, the switch is likely to have been completed before the called party picks up the phone. In such a case no PCM packets will be sent at all.

In the case where trade-off negotiations between packet size and rate for a constant encoding rate are taking place, there is no guarantee that the Flow-Spec negotiation will terminate in a clear success or failure. The ORIGIN PVT is following a trial-and-error strategy in a situation where the network conditions are varying with the actions of others. Oscillation is likely either due to load effects or conflicting limitations arising in different nets along the route. For example, a call across the WB SATNET to a Packet Radio (PR) net might find that the SATNET could handle the flow if many small packets were used but that the PR net could cope only if the flow involved fewer larger packets. Negotiation under such conditions will result in oscillation unless the PVTs and Gateways take steps to

prevent it. In our implementation, the ORIGIN PVT uses the size and peak rate limits provided by the Gateway in the Flow-Spec and changes its request in the direction indicated by the limitation bits until either the relevant network limit is reached or some internal PVT limit is encountered. For example, if the limitation bits indicate that the peak rate is too high, the PVT will increase the packet size to the maximum value, thereby minimizing the rate. If this request fails, the PVT will give up and close the connection. This strategy prevents oscillation but can cause some calls to fail that might succeed with a more complex search algorithm. Currently, the Gateways will close a connection if the ORIGIN does not respond with a new CONNECT after being sent an ACCEPT with limitation bits indicating a need for further negotiation. We expect to add another mechanism to close a connection that has remained in an unsatisfied state for too long a time. This new mechanism would allow the PVTs to use search strategies that might oscillate without having to be concerned with termination criteria.

The negotiations described here provide for adjusting flows to match network capacities, but the mechanisms operate automatically only in the direction of reducing the capacities available to individual connections. The Gateways do not remember the fact that a connection started out requesting a higher rate and was subsequently cut back due to other traffic. As a result, there is no mechanism for automatically returning to higher rates when the other traffic goes away. In our implementation, the only way for a user to get back to a higher rate is to initiate the request by actuating the encoding rate switch on the PVT. Pushing the switch causes the PVT to attempt a change to the requested rate. The process involves the sending of a new CONNECT message by the ORIGIN PVT and will result in a change to a higher rate if the capacity is available at the time the switch is activated.

5. Flow-Spec Experiments

A number of experiments/demonstrations have been carried out to show the capabilities of the Gateways in managing network resources and of the PVTs in adapting to varying network conditions. The experiments have made use of PVTs with two encoding rates, PCM at 64 kbps and LPC at 2.4 kbps. In one demonstration scenario, we set the WB SATNET stream capacity to be just large enough for one PCM call. We then place such a call, which goes through successfully, and attempt a second one. The Gateways discover that there are not enough resources when the PSAT refuses to increase the stream capacity. They then look for connections that are not running at minimum rates and demand changes to lower rates by sending ACCEPTs to the originating PVTs with appropriate Flow-Spec limitation bits set. The connections are attacked one at a time until the total flow requirements of the connections are compatible with the stream capacity. In this scenario, both calls end up at LPC rates.

In a second scenario, we set the Gateway flow limits for one of our LEXNETs to values appropriate for a Packet Radio net, i.e., packet size limit of 128 words and packet rate limit of 10 packets per second. We then attempt to set up a call from a regular

LEXNET PVT to one on the simulated PR net. The call starts out requesting PCM rates, gets switched to LPC because the simulated PR net cannot handle PCM call rates, and then goes through a trade-off negotiation to get the packet rate equal to or less than the 10 pps limit for the simulated PR net. (The normal LEXNET LPC encoding runs at 25 packets per second.) In the case of a real PR net, the packet voice terminal on that net would know that it could not handle PCM rates, and the rates, and the initial NVP call setup would have selected LPC or some other low-rate encoder prior to setting up the ST connection. The packet size/rate tradeoff would still have been negotiated via the ST Flow-Spec as in our scenario.

6. Flow-Spec for Conferences

In the case of conferencing, the Flow-Spec specifies the overall characteristics of the conference (i.e., packet size and peak packet rate) rather than the needs of an individual speaker. Because of its global nature, the conference Flow-Spec is defined by the Conference Access Controller (CAC) for each conference depending on the type of voice encoding to be used in the conference. When a Gateway becomes involved in supporting a conference, it asks the CAC for information about the conference. A part of the response is the conference Flow-Spec which the Gateway uses to allocate resources for the conference and to free the appropriate resources when the conference ends. Because the conference setup proceeds in stages as new participants join in, negotiation between the conference and the Gateways about flow requirements would be very complex. We chose to avoid such complexity by treating the global Flow-Spec as non-negotiable with the result that a particular conferee will be able to join a conference only if resources are available to support his communications at the globally specified rates. For the same reason, we do not provide for a conference to change encoding rates after it has been established. The Flow-Spec for a conference, therefore, has only two parameters: packet size and peak packet rate. Limitation bits, etc., are not needed.

7. Current Status

Gateways supporting the Flow-Spec option for point-to-point connections are operating at the ISI, SRI, DCEC, and Lincoln sites. Packet voice terminal (PVT) programs to use the Flow-Spec for negotiating with the Gateways for resource allocation are operating in the PVTs at Lincoln Laboratory. The corresponding code for the Switched Telephone Network Interface (STNI) PVTs is ready, but has not yet been installed. Checkout is under way on the extension of the Flow-Spec for conferences.

D. GATEWAY DEVELOPMENTS

Extensions were made to the Gateway software to provide conference shutdown, shutdown of 'stuck' point-to-point connections, forwarding to 'unknown' networks, choice of slotsize for the nominal WB SATNET stream, and asynchronous terminal output.

1. Conference Shutdown

The design of the Gateway's internal data structures for maintaining conference information involves compact bit maps to represent the participants in a conference and to provide efficient forwarding of packets using broadcast addresses. A deficiency of the bit maps is the unavailability of the full IP addresses of the participants and any knowledge about the presence of connections between any two participants. Therefore, when one participant disconnects from another, the Gateway cannot easily determine whether the participant has thus disconnected from all other participants and can thereby be removed from the conference. On a larger scale, the Gateway cannot easily determine whether all participants have left the conference and the conference can thereby be terminated.

Until now, Gateway support for conferences involved manual operations, both for allocating network resources and for eventually shutting down the conference. With the advent of Flow-Spec, network resources now are allocated automatically by the Gateway when a conference opens, such resources not being released until the conference shuts down. These resources include expensive stream capacity if the conference involves the WB SATNET. Hence, the need for the Gateway to determine automatically when to shut down a conference and reclaim its resources has gained greater importance.

Two independent mechanisms were devised for determining when to terminate a conference. The first scheme involves querying the Conference Access Controller whenever the Gateway processes an ST DISCONNECT or REFUSE indicating that the participant possibly is leaving the conference. Since part of the protocol between conference participants and the Access Controller involves messages about entering and leaving the conference, the Access Controller potentially knows who is currently in the conference. If the Access Controller tells the Gateway that all participants have left, then the Gateway shuts down the conference and releases its resources. The Access Controller, however, may not know that all participants have left, since intermediate network breakage may have prevented a participant from communicating that he is leaving. Therefore, as a backup scheme, the Gateway monitors the arrival of conference speech packets. If such packets do not arrive for a long period of time, the assumption is made that the conference has terminated and the Gateway shuts it down and releases its resources.

2. Shutdown of Point-to-Point Connections

A point-to-point connection may not close normally for several reasons. Two prominent examples are breakage of an intermediate network preventing an ST DISCONNECT or REFUSE from getting through and, in the case of an STNI call to a time/weather service, failure by the human user to hang up the STNI (via the '*#' sequence) prior to physically hanging up the phone. Two mechanisms were implemented to provide for the termination of such 'stuck' connections. In the first mechanism, the Gateway does not give up when it times out an attempted ST protocol shutdown sequence. Instead, it keeps trying indefinitely at infrequent intervals thereby providing for recovery after a temporary outage of an

intermediate network. The second mechanism is a new command that may be typed to the Gateway to instruct it to initiate the ST point-to-point shutdown protocol. An STNI PVT can be unstuck by use of this new command.

3. Forwarding to 'Unknown' Networks

To accommodate packets from/to networks about which the ST Gateway does not know routing information, the ST Gateway now forwards such packets to one of a set of Internet IP Gateways for eventual forwarding to the ultimate network. In some cases, this may first involve forwarding the packet from one ST Gateway to another. If the packets being forwarded are ST packets, they are encapsulated in IP messages before being sent through an IP Gateway that does not understand the ST protocol.

4. Nominal Stream Slotsize

When the Gateway receives an ST Connect for a connection that will require transmission through the WB SATNET, two actions must be performed: the ST Connect (and subsequent ST control messages) must be forwarded, and stream slotsize must be reserved for the connection. Two questions arise. Should the ST control messages be forwarded via datagrams or via a stream? Should the stream slotsize be reserved before the ST control messages are forwarded or should the two actions be performed in parallel?

In the Gateway design, we chose stream rather than datagram transmission for forwarding the ST control messages. This choice was dictated by the observed problem of PSAT crashes when too many datagrams are being transmitted from all sources in the WB SATNET. We also chose to maintain at all times a stream of nominal slotsize for the forwarding of the ST control messages and to transmit such messages in parallel with the request to the PSAT to enlarge the stream slotsize. This was motivated by our desire to provide speedy establishment of connections and by the observed length of time (on the order of a second or two) that it takes to reserve or change stream slotsize.

We therefore need to determine a reasonable slotsize for the Gateway's nominal stream. A slotsize of 40 words has been determined to be sufficient for the largest IP or ST message that possibly could occur in the establishment of a connection. We also have reordered the output queues to favor the transmission of ST control messages. The choice of the proper slotsize together with the queue reordering has resulted in speedy and reliable establishment of ST connections at the price of maintaining a small nominal stream.

Since the ST control messages are being exchanged in parallel with the stream enlargement request, the question then remains as to what to do with speech packets that may arrive if the connection is established before the stream is enlarged. Since there are likely to be few such packets, we decided to allow them to be discarded if the length of time that they are queued prevents their timely transmission.

5. Asynchronous Terminal Output

In order to support remote access of Gateways via error-correcting devices attached to modems, the Gateway's logging of status and error information was changed from synchronous to asynchronous typeouts. This change is transparent to most Gateway usage, but prevents the Gateway from being held up due to a disconnection of the telephone line between the computer on which the Gateway is running and the terminal on which the logging is taking place. In the case of a long outage, logging or error information may be lost but the mainline gateway functions of managing connections and forwarding packets continue without interruption. Provision is made for on-line usage where the Gateway is typing out on a slow terminal.

E. SUN WORKSTATIONS

We are still exploring the possibilities of using the SPP to provide voice as an added capability to SUN workstations connected on a local computer network. Thus far, we have not been able to avoid the bottleneck presented by the UNIX operating system in the stations. Even by carefully optimizing the buffer size parameter within the I/O routines, we were unable to support a continuous full-duplex loop between the SPP and the SUN workstation.

We learned that SRI was trying to develop a similar system, on another DoD-sponsored program, using the CHI-5 voice processor instead of the SPP. They have a newer version of the workstation hardware and also have developed contacts with the people at SUN. Since SRI was having a problem with the flow control on the CHI-5, and since one of their goals also is to achieve full-duplex speech through the SUN, we lent them an SPP in late May. They have interfaced the SPP to their SUN workstation and have been attempting to solve the real-time problems with some support from SUN. We are providing support to SRI on the use of the SPP and are tracking their progress.

The status at the end of September was that SRI had developed a new real-time-oriented I/O Driver for the UNIX operating system in the SUN, as an approach to overcoming the problems encountered with real-time throughput in UNIX. The new driver had been configured into the UNIX kernel, but had not yet been tested. SRI planned to continue working on the UNIX I/O problem in cooperation with the people at SUN.

F. ETHERNET PVT

The development of an Ethernet Concentrator Interface (ECI) for replacing its LEXNET counterpart in either a LEXNET Concentrator Interface (LCI) or Packet Voice Terminal (PVT) is concluding. The new Ethernet Concentrator Interface is packaged on two PVT-compatible wirewrap circuit cards, functionally partitioned into CPU and I/O sections.

Since the last semiannual report, a preliminary prototype ECI has been built and tested, followed by a final choice of Ethernet support chips and layout of CPU and I/O boards as shown in Figure 5.

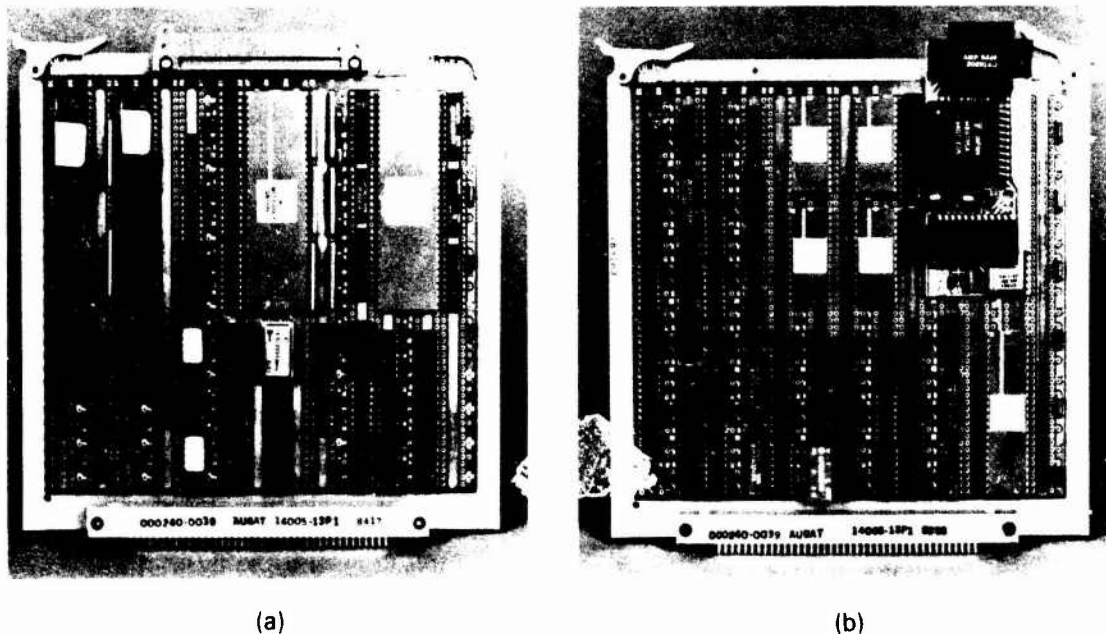


Figure 5. (a) ECI CPU board. (b) ECI I/O board.

The ECI CPU board now contains the 68000 CPU, 68450 DMAC, 32K bytes each of EPROM and RAM, plus overhead logic. The I/O board contains a 5 byte-wide Direct Memory Access (DMA) port, a 68901 multifunction chip used as a parallel port and interrupt controller, a printed circuit subsystem card for the Ethernet Data Link Controller (EDLC) and Manchester codec, along with FIFO buffers and support circuitry.

As reported earlier, several teams of integrated circuit makers are developing Ethernet VLSI support chips. However, only Fujitsu chips were available for our immediate use, and required additional interface design and packaging effort. A small printed circuit subsystem board has been designed to package the data link controller, Manchester codec, oscillator, discrete components, and transceiver cable connector. This PC board plugs into the IC lead sockets of the I/O wirewrap card.

To attain the full-duplex Ethernet data rates (20 Mbits/s) required for local loopback testing under DMA burst-mode control, 128 bytes of transmit and receive FIFO buffer memory have been included. Automatic starting and stopping of DMA activity is now possible while providing overflow/underflow protection at the EDLC.

Operational software has been completed during the past quarter that should permit the orderly substitution of ECI CPU and I/O cards for the Buffer Control and Modem cards currently used within LEXNET terminals. Ancillary programs are also available for echoing packets between terminal pairs on a private coaxial cable. Additionally, a number of concise assembly language routines are on hand as a by-product of tests conducted on memory, DMA, interrupts, and I/O ports.

Final system tests are currently in progress.

III. VOICE-CONTROLLED SYSTEMS

A. ROBUST SPEECH RECOGNITION OVERVIEW

Work on robust speech recognition has led to off-line recognition systems written in 'C' than run on a VAX computer using array processors and to interactive systems that run on a high-speed LDSP speech processor attached to a PDP-11/44. Off-line systems have been used to develop a new noise-resistant open-endpoint algorithm, to develop a new spectral tilt pre-emphasis technique, to develop network-based recognition systems, and to design algorithms to be incorporated in the VLSI DTW wafer. Interactive systems have been used to develop and test a new spectrally-based open-endpoint recognition system.

A general purpose off-line recognition system was used to test both a noise-resistant open-endpoint DTW algorithm and frame-by-frame spectral-tilt pre-emphasis that normalizes the spectral tilt of each speech frame. This pre-emphasis was added to reduce the effect of spectral tilt on recognition to better model human speech perception. It also tends to reduce the effects of stress and variations in the glottal source on recognition accuracy. Experiments were performed using a 25-word vocabulary spoken by one talker in the AFTI speech data base and recorded in noise levels ranging from 85 to 112 dBA. Recognition accuracy was tested using the open-endpoint DTW algorithm with LPC parameter extraction with and without spectral tilt pre-emphasis. Accuracy was higher with spectral tilt pre-emphasis, and the increase in performance with fixed pre-emphasis ranged from 2.5 to 5.5 percentage points.

A recognition system has been developed on the Lincoln Digital Signal Processor (LDSP) fast signal processing computer that uses an open-endpoint spectrally-based DTW recognition algorithm. Parameter extraction in this system attempts to model some aspects of human hearing by using adaptive critical band filters extracted from a high-resolution spectrum. Recognition accuracy with one female talker and three repetitions of the digits 0 to 9 was 100%. The system currently is being tested with the Advanced Fighter Technology Integrator (AFTI) data base.

Off-line recognition systems have been used to explore network-based recognition structures. One structure provided a new and more accurate model of the duration of speech events, compensated carefully for additive noise in the distance metric, used noise anchors at the beginning and end of word models, and included pruning to reduce computation time. A new training technique using K-means clustering was developed to build word models automatically in this network structure. Although the performance of this system was good, it was computationally too costly (25 MIPs for an isolated-digit recognition system) and required large turnaround time for small experiments. Attention was switched to Hidden Markov Model (HMM) systems that are more computationally efficient and have been shown elsewhere to perform well. Both Viterbi and Full Probability decoders were written, and then a new training procedure for HMM systems was developed that

attempts to select parameters for HMM word models that are globally optimum for the training data. This procedure uses a technique called simulated annealing to select model parameters. It currently is being tested using Monte Carlo techniques.

An off-line recognition system is being used to guide trade-off decisions that must be made in designing a VLSI DTW wafer and the recognition system in which the wafer will be embedded. Both isolated-word and connected-word speech data bases have been run through the system to evaluate different hardware designs.

B. OPEN-ENDPOINT DYNAMIC TIME WARPING

Robust endpoint detection is a crucial and difficult problem in noisy speech recognition. We have been approaching this problem by applying an open-endpoint DTW algorithm that does not require explicit determination of speech endpoints prior to the recognition process. The endpoint detection is embedded in the DTW process, and could be used with either a spectral or an LPC front end. We will describe briefly the open-endpoint DTW algorithm here.

The typical shortest-path dynamic programming problem is a global optimization problem, which may be embedded in a set of local optimization problems, one for each node, and solved by recursively fixing individual nodes, from the beginning node to the terminating node. The local optimization problems may be described by the following equation:

$$D_{ij} = d_{ij} + \min_{\substack{\text{allowed} \\ k, l}} D_{kl} w_{kl,ij} \quad (1)$$

where d_{ij} is the local distance measure at the node (i, j) , D_{ij} is the cumulative distance measure, and $w_{kl,ij}$ is the weight put on by taking the transition from the node (k, l) to the node (i, j) .

When applying this method to speech pattern recognition, it is typically necessary to identify the beginning and ending instances of words. Since normal speech contains silent periods, nonvoice noisy periods (microphone clicks, background noise, etc.), and nonspeech voice periods (lip smacks, breath noise, etc.), robust endpoint detection is difficult, and endpoint errors can cause recognition errors.

A different view of the dynamic programming procedure provides an elegant alternative to the above approach. We arrange test speech and reference template on a two-dimensional plane with input speech on the positive x axis and reference template on the negative y axis. We apply Equation (1) to each individual node, assuming all D_{kl} exist and all boundary conditions are taken care of. For a given node (i, j) , a unique node (k, l) (assume all nodes have different cumulative distances) can be identified through Equation (1), hence a unique path can be found to the upper left quadrant of node (i, j) .

When the path is traced back to the x axis (corresponds to the beginning of reference template), the beginning of the word has been identified, given that (i, j) is a node on the path. When j corresponds to the end of the reference, we then have a unique path corresponding to the given reference, and a cumulative distance measure that reflects the goodness of the path.

A word can be spotted by comparing the cumulative distance measure with a threshold, and the spotted word is a hypothesis. Ambiguous hypotheses must be resolved by a high-level decision mechanism.

The computation of Equation (1) is simply structured, so the method allows parallel computation. In sequential machines, it is straightforward to prune impossible paths and reduce computation considerably. Our experience shows that a simple pruning method with a conservative pruning threshold can reduce the computation of d_{ij} (by far the most costly computation) by 70% for isolated English digits, recorded with about 50% silence intervals. The pruning algorithm is suggested by Equation (1). If all D_{kl} for all allowed k and l are greater than a threshold, then the local distance d_{ij} need not be computed since the optimal path will never go through this node. The risk of erroneously pruning off a legal path can be reduced to a minimum if the threshold is chosen conservatively; on the other hand, a tighter threshold will be able to differentiate local differences such as that presented by 'bee' and 'gee'. The trade-offs will be studied by extensive experiments.

Figure 6 illustrates the advantage of open-endpoint DTW versus DTW with endpoint detection. Figure 6(a) shows the energy profile of a segment of input speech that contains a

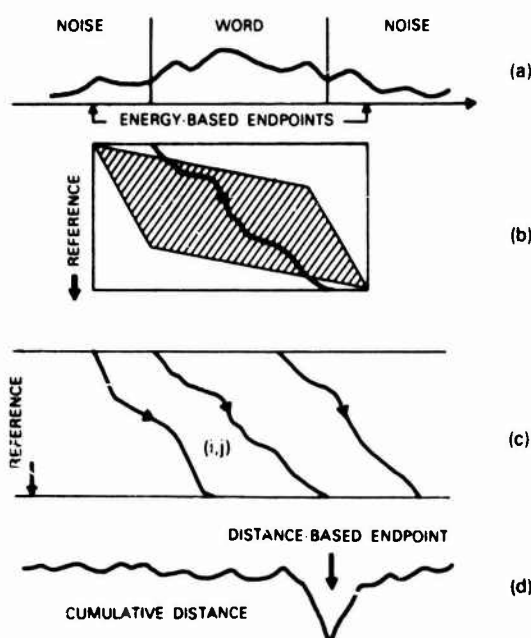


Figure 6. Open-endpoint template matching.

word and surrounding noises. The noise may be background noise, breath noise, lip smack, etc. An energy-based endpoint detector probably will include noise segments as well as the word, hence the DTW comparison to reference template does not contain the true path, as shown in Figure 6(b). The open-endpoint DTW, however, makes no *a priori* assumption about word beginning and ending; therefore the optimal match is found by the more reliable spectral distance measure, as shown in Figure 6(c) and Figure 6(d).

C. DTW-BASED NOISY SPEECH EXPERIMENTS

1. Off-Line Recognition System

An off-line, highly-modularized recognition system has been written in the 'C' language on a VAX computer, to be used for algorithm development and refinement. The off-line system includes LPC and Spectral Envelope Estimation (SEE)¹ front ends, training and recognition modes, and the flexibility to test a variety of recognition algorithms. Computation in the system takes advantage of an AP-120B array processor, connected as a peripheral to the VAX. Specifically, LPC distance computations are currently carried out on the array processor. The use of the array processor facilitates extensive testing in the multiuser VAX environment, since array processor computations can proceed in parallel with other computations on the VAX.

2. Experiments With Variable Pre-emphasis

The off-line recognition system has been used for experiments on speech data from the AFTI F16 data base. The system as tested included: open-endpoint DTW with LPC parameter extraction, variable frame-by-frame spectral tilt pre-emphasis, and a transformed Itakura distance metric designed to reduce insertion and substitution errors by amplifying local large errors. As detailed below, the variable spectral tilt pre-emphasis provided enhanced accuracy over fixed pre-emphasis.

Perception experiments have indicated that humans are remarkably insensitive to changes in the depth of spectrum valleys, in the overall spectral tilt, and in relative formant amplitude. The widely used Itakura-Saito maximum likelihood ratio LPC distance measure agrees with those properties of human speech perception in that it is more sensitive to spectral peaks than to valleys; however, the Itakura-Saito measure is also quite sensitive to spectral tilt and to relative formant amplitudes.

We have been investigating the idea of prefiltering speech with a variable one-pole inverse filter for application in speech recognition. In the variable pre-emphasis technique, a first-order inverse LPC filter is computed for each frame and used as the pre-emphasis filter. This reduces the overall spectrum tilt that often is caused by such factors as ambient noise, talker stress, or distance of the talker from the microphone. Inverse filtering is applied to both training and test data for the recognition application.

We also have modified the Itakura-Saito distance measure using the following transformation:

$$f(x) = x \quad 1 < x \leq a$$

$$f(x) = c(x-a) + a \quad x > a$$

where x is the Itakura-Saito distance. With $c > 1$, this transformation effectively amplifies large errors; we typically have been using $c = 2$, $a = 5$. The purpose of this transformation is to reduce insertion and substitution errors by paying more attention to local variances. For example, the substitution error that sometimes occurs between the words 'degrees' and 'east' is reduced significantly. The transformation also seems to reduce insertion errors, although more tests need to be done to determine its general effectiveness.

The results of preliminary experiments with this LPC/DTW system, using a 25-word vocabulary spoken by one talker in the AFTI speech data base, are shown in Figure 7. Tests at noise levels from 85 to 112 dBA were conducted, with both fixed and variable pre-emphasis. The references were generated from the first occurrence of each vocabulary word in the recording at the 85 dBA noise level. The variable pre-emphasis technique performs as well as or better than fixed pre-emphasis at all noise levels. The 5% error rate improvement with variable pre-emphasis at the highest noise level is a particularly encouraging result.

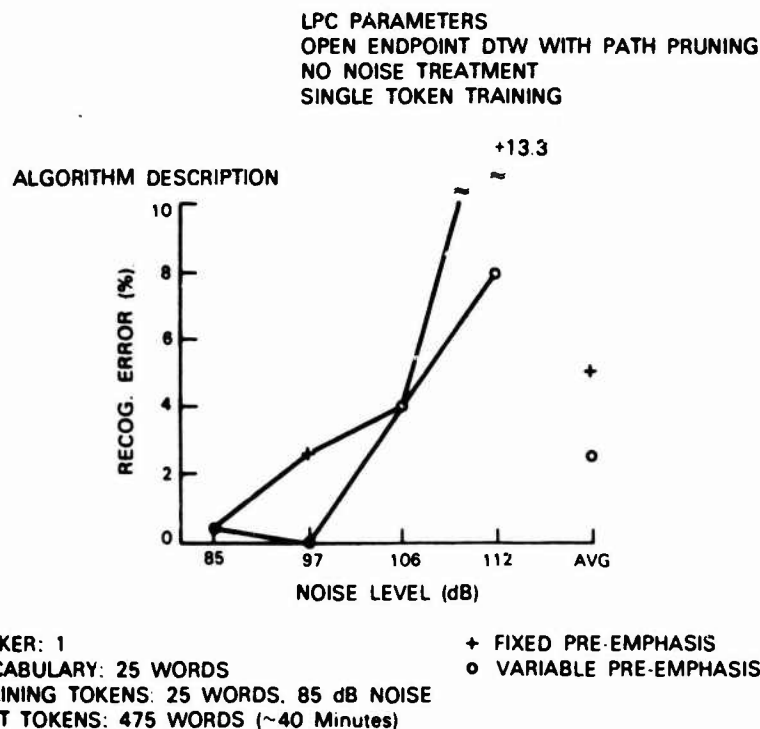


Figure 7. Preliminary processing results on AFTI data.

D. LDSP-BASED NOISY SPEECH RECOGNITION SYSTEM

An experimental recognition facility has been implemented using two LDSPs operating in conjunction with our speech PDP-11 computer. The system uses open-endpoint DTW and includes a new spectrally-based front end designed to achieve robustness by focusing on spectral peaks. This front end, as described below, models aspects of human hearing by using adaptive critical band filters extracted from a high resolution spectrum. Initial recognition experiments have yielded promising results.

Probably the most common approach to frequency domain speech recognition is to use a bank of critical band filters. It often is argued that since the peripheral auditory system can be modeled by such a filter bank, then the minimal set of perceptual units are the downsampled envelope-detected outputs of these filters. Whenever such a filter bank has been used for recognition in the past, a fixed set of a relatively few (19-30) filters has been used. The problems with such a filter bank are well known, the main criticism being the sensitivity to pitch which renders the spectral templates rather unstable. Of course the ear avoids these problems by using a set of some 30,000 filters. However, over any one analysis interval, most of the speech energy is located within small regions about the peaks of the high-resolution spectrum. Therefore, one can imagine that of the 30,000 possible critical band filters, the important information will be contained in the subset of filters located at each of the resolved peaks. Therefore, we have attempted to model this process by using an adaptive set of critical band filters by centering a filter of the appropriate bandwidth at the frequency of each of the resolved peaks. The total power out of each filter is taken to be the sum of the powers contributed by all of the peaks lying within the 40 dB edges of that filter weighted by a 3rd order Butterworth gain characteristic. One can imagine that the 30,000 filters in the ear allow for the computation of a smooth critical-band spectral envelope (not to be confused with the vocal tract spectral envelope; they are quite different). We can approximate this envelope simply by linearly interpolating between the power measurements made at each of the resolved peaks.

One of the major arguments in favor of the critical band filters is that an orthogonal set of such filters spans the perceptual measurement space. To model this, we define such an orthogonal set using 31 filters with bandwidths increasing from 100 Hz at 200 Hz, up to 900 Hz at 4500 Hz. The outputs of such a filter bank are estimated by sampling the critical band envelope that was generated from the high-resolution spectrum at the logarithmically increasing critical band frequencies. These 31 power measurements then become the input to the speech recognition algorithm.

In preparation for processing analog speech data (including the AFTI data) using this system, the data was digitized with the aid of an interactive display that assured that each word, and any surrounding breath noise, was included in a 3-second interval of speech. The resulting speech then was processed and reduced to 31 power measurements per 10-ms frame and stored on disk. Because of the heavy interaction between the LDSPs and the PDP-11, this stage of the system is very time-consuming.

The recognition stage involves an LDSP with large outboard memory and a file-handling program in the PDP-11. The LDSP receives an input segment of parameters representing three seconds of speech from the PDP-11 and stores this in outboard memory. It then receives one reference segment at a time and performs the open-endpoint DTW. A straightforward squared dB error criterion currently is used as the distance measure. The score of the best path along with the corresponding beginning and end frames of the input segment are sent back to the PDP-11.

Initial experimentation was done on the digits (0 through 9) for a female speaker (pitch 170-200 Hz), with three repetitions of each digit. No errors were made. Equally important is the fact that the 900-word comparisons were processed in two hours, and this includes considerable interaction between the LDSPs and the PDP-11. Evaluation of the performance of this recognition system then began using a single speaker from the AFTI data base. Tests for the 85-dBA SPL have been completed and the matches all have been correct. Tests for the lower signal-to-noise ratio cases are underway.

E. NETWORK-BASED SPEECH RECOGNITION SYSTEM DEVELOPMENT

We have performed some experiments with a new network-based speech recognition structure. The basic internode path topology was not new, but the nodal structure itself was new. A path incurred several costs when it traversed a node. The first cost was a spectral cost computed by a spectral distance measure. Initial experiments assessed the spectral cost at node entry time and the later experiments used the average spectral distance during node residence as the spectral cost. The second cost was a node residence-time-based cost function. The total path cost was the sum of the individual costs at all nodes on the path.

The distance functions tested were non-negative slope transforms of a perceptually motivated, adaptively weighted L2 norm of the difference of energy-normalized log spectra². The three transforms were square, linear, and an approximation to the Itakura-Saito log-likelihood ratio (LLR)³. The square gave the best results. Adaptive background templates were used to allow for variations in the acoustic environment.

The node residence time cost was the negative log of a Gaussian fit to the training data (which results in a parabola). As more data was entrained, it was expected that better residence cost functions could be generated from a histogram of past residence times using a Viterbi decoding scheme.

A number of training schemes were tried. The first scheme moved to the next node whenever the spectral distance from the current node exceeded a threshold. If the next node(s) did not exist or also exceeded the spectral distance threshold, a new node was created. Another set of training schemes generated a network implementation of a modified DTW with a decimated template for decimation factors of 2 to 5. The final and most promising scheme used a K-means-like algorithm to assign zones to the training words from which the spectra and residence times were taken.

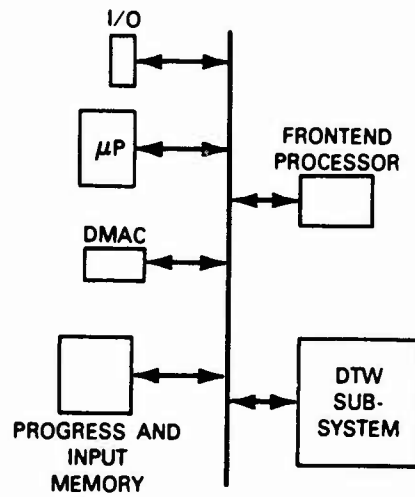


Figure 8. ASRU system block diagram.

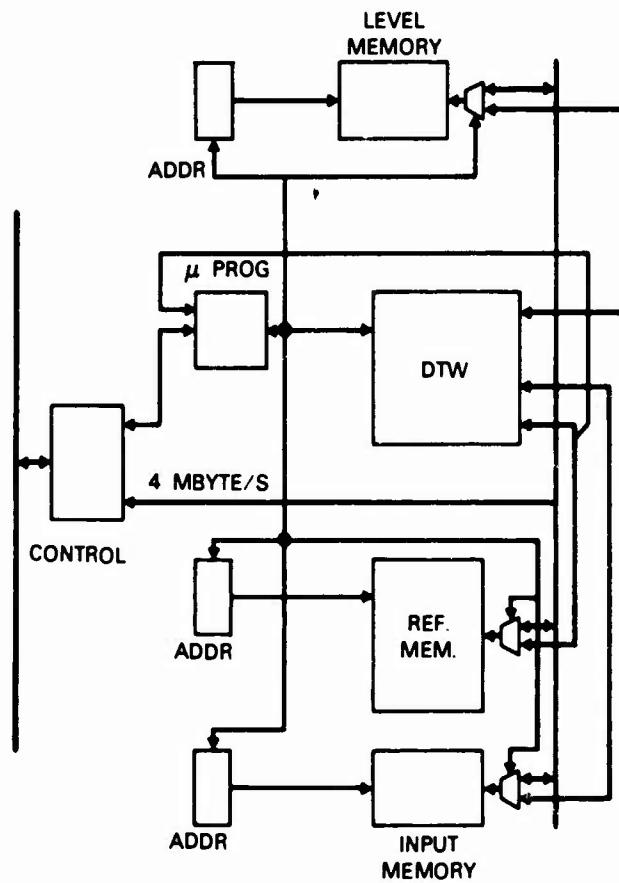


Figure 9. DTW subsystem.

The analyzer used a process-based implementation. A number of path-pruning algorithms were used. The first was a fixed threshold as a function of path length strategy. Adaptive threshold strategies also were tested. The best appeared to be a 'best N' path strategy.

Initial tests were performed on a small isolated digits data base. Since a word in this data base is surrounded by background noise, a few frames of background noise at the beginning and end of the word (left and right 'anchors') were matched to improve the (open) endpoint location. The recognition accuracy was promising, but the quantity of computation (25 MIPs for the best pruning scheme) required was very large for such a small problem. This led to very large experimental turnaround times, which would have become prohibitive for larger vocabularies and/or continuous speech.

Hidden Markov model (HMM) systems are far more computationally efficient during recognition, but have a weaker speech model. HMM systems also have shown the best performance to date for large vocabulary continuous speech recognition. HMM systems are trained from observed speech data, but the methods for training are gradient-like methods that converge to a local minimum. It was realized that a new stochastic optimization method, simulated annealing⁴, should be able to train the HMM recognizer with a high probability of finding the global minimum.

The immediate goal of this work is to demonstrate the training method using digital Markov sources. Current results are favorable, but more needs to be done to prove the method. Once enough has been learned using the digital sources, the simulated annealing will be used to train a recognizer so that recognition performance comparisons can be made between the new method of training and the standard methods.

F. ADVANCED SPEECH RESOURCE UNIT SIMULATION FACILITY

Lincoln's DARPA-sponsored Restructurable VLSI Program includes the development of a wafer-scale Dynamic Time Warping device for isolated-word and connected-word speech recognition. To operate as a speech-recognition system, the DTW device must be supported by appropriate data-processing and voice-processing subsystems. We refer to the composite system as an Advanced Speech Resource Unit (ASRU), which we expect will be composed of a single board including DTW template matching, voice processing, and data processing.

As part of the Wideband Program, we have implemented software simulations of the voice-processing and data-processing portions of the ASRU to provide support for the design and test of the DTW wafer. We also have developed high-level block diagram designs for the ASRU, as shown in Figures 8 and 9. The overall system can be described as a general purpose processor with front end and word matching peripheral processors. The DTW subsystem performs the level-building speech recognition algorithm. Its key component is the VLSI DTW wafer.

The voice-processing front-end now is expected to be a filter bank due to the ability to represent spectral parameters with fewer bits than in an LPC representation. This simplifies

the internal DTW architecture and wafer/host communication. The DTW subsystem is essentially a microprogram-controlled peripheral processor whose input data and control commands are provided by the host system by permitting word parallel operations.

A high-level simulation package, written in the 'C' language for the VAX computer, has been developed and utilized to assist in the design decision process concerning both the internal DTW wafer architecture and high-level system considerations. The software performs simulation tests on a fixed data base by continuously varying system parameters while automatic analysis maintains statistics on the recognition performance. These results then may be displayed or plotted for easy analysis. Included in the package of simulation software are: level-building DTW template matching, an LPC front end, a filter bank front end with automatic gain control and pre-emphasis options, a down-sampling mechanism and documentation and analysis tools.

During the last quarter, several enhancements to the software were made in the areas of expanded system options and in documentation and analysis. A full set of path constraints and weightings are now available. Statistics now are maintained on the selected path accumulation values. This allows the observance of the range of values encountered in the matching process as well as specifying limits on internal word sizes. The ability to produce hardcopy output of the screen documentation has been added to assist in recording performance results with as little effort as possible. Added to the screen display is the system and data base description for easy identification and categorization.

Tests are being run on two standard data bases: TI's 16-Speaker Isolated Word Database and TI's Connected Digit Database. A final decision has not yet been made on the detailed parameters of the DTW architecture, but operating regions have been established and are under thorough investigation.

Work on the DTW wafer and on its demonstration in an ASRU system structure will continue in FY85 under the DARPA Restructurable VLSI Program. That effort will include: final definition (aided by simulation) of the DTW parameters, and detailed design and development of the ASRU demonstration system; as well as layout and fabrication of the DTW wafer itself.

REFERENCES

1. D.B. Paul, "The Spectral Envelope Estimation Vocoder," IEEE Trans. ASSP, **29**, No. 4, 786-793, (August 1981).
2. D.B. Paul, "An 800 BPS Adaptive Vector Quantization Vocoder Using a Perceptual Distance Measure," ICASSP '83, Boston, Massachusetts, (April 1983), **14**, 73-76.
3. S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, "Optimization by Simulated Annealing," Science, **220**, No. 4598, (13 May 1983).
4. A.H. Gray, Jr., J.D. Markel, "Distance Measures for Speech Processing," IEEE Trans. ASSP, **24**, No. 5, 380-391, (October 1976).

GLOSSARY

AFTI	Advanced Fighter Technology Integrator
AR	Adams-Russell Company
ASRU	Advanced Speech Resource Unit
BBN	Bolt, Beranek and Newman
CAC	Conference Access Controller
DCEC	Defense Communications Engineering Center
DMA	Direct Memory Access
DRT	Diagnostic Rhyme Test
DTW	Dynamic Time Warping
ECI	Ethernet Concentrator Interface
EDLC	Ethernet Data Link Controller
EISN	Experimental Integrated Switched Network
ESI	Earth Station Interface
GTG	Gateway-to-Gateway
HMM	Hidden Markov Model
ICMP	Internet Control Message Protocol
IP	Internet Protocol
ISI	Information Sciences Institute
LCI	LEXNET Concentrator Interface
LDSP	Lincoln Digital Signal Processor
LEXNET	Lincoln Experimental Packet Voice Network
LLR	Log-Likelihood Ratio
LPC	Linear Predictive Coding
MH	Measurement Host
MPP	Multipurpose Packets
NOC	Network Operations Center
NVP	Network Voice Protocol
PC	Printed Circuit
PCI	Packet/Circuit Interface
PCM	Pulse Code Modulation
PR	Packet Radio
PSAT	Packet Satellite Interface Message Processor
PVT	Packet Voice Terminal

SEE	Spectral Envelope Estimation
SPP	Speech-Processing Peripheral
SRI	SRI International
ST	Stream Protocol
STNI	Switched Telephone Network Interface
TCP	Transmission Control Protocol
WB SATNET	Wideband Satellite Network

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM												
1. REPORT NUMBER ESD-TR-84-284	2. GOVT ACCESSION NO. AD-A152 327	3. RECIPIENT'S CATALOG NUMBER												
4. TITLE (and Subtitle) Wideband Integrated Voice/Data Technology		5. TYPE OF REPORT & PERIOD COVERED Semiannual Technical Summary 1 April — 30 September 1984												
		6. PERFORMING ORG. REPORT NUMBER												
7. AUTHOR(s) Clifford J. Weinstein		8. CONTRACT OR GRANT NUMBER(s) F19628-85-C-0002												
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lincoln Laboratory, M.I.T. P.O. Box 73 Lexington, MA 02173-0073		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element No. 62708E Project No. 3T10 ARPA Order 3673												
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, VA 22209		12. REPORT DATE 30 September 1984												
		13. NUMBER OF PAGES 52												
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Electronic Systems Division Hanscom AFB, MA 01731		15. SECURITY CLASS. (of this report) Unclassified												
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE												
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.														
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)														
18. SUPPLEMENTARY NOTES None														
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)														
<table border="0"> <tr> <td>packet speech ,</td> <td>internetwork ,</td> <td>speech recognition,</td> </tr> <tr> <td>Linear Predictive Coding ,</td> <td>protocol ,</td> <td>voice conferencing ,</td> </tr> <tr> <td>LEXNET ,</td> <td>gateway ,</td> <td>packet voice terminal ,</td> </tr> <tr> <td>Wideband SATNET</td> <td>ARPANET</td> <td>Dynamic Time Warping</td> </tr> </table>			packet speech ,	internetwork ,	speech recognition,	Linear Predictive Coding ,	protocol ,	voice conferencing ,	LEXNET ,	gateway ,	packet voice terminal ,	Wideband SATNET	ARPANET	Dynamic Time Warping
packet speech ,	internetwork ,	speech recognition,												
Linear Predictive Coding ,	protocol ,	voice conferencing ,												
LEXNET ,	gateway ,	packet voice terminal ,												
Wideband SATNET	ARPANET	Dynamic Time Warping												
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)														
<p>This report describes work performed on the Wideband Integrated Voice/Data Technology Program sponsored by the Information Processing Techniques Office of the Defense Advanced Research Projects Agency during the period 1 April through 30 September 1984.</p> <p>ORIGINATOR - SUPPLIED KEY WORDS INCLUDE:</p>														

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)